Giustino Gaspa, Roel F. Veerkamp, Mario P.L. Calus, Jack J. Windig,

Assessment of genomic selection for introgression of polledness into Holstein Friesian cattle by simulation,

Livestock Science,Volume 179,2015, 86-95,
https://doi.org/10.1016/j.livsci.2015.05.020.

Assessment of Genomic Selection for introgression of polledness into Holstein Friesian cattle by simulation.

Giustino Gaspa[a,b], Roel F. Veerkamp[a], Mario P. L. Calus[a], Jack J. Windig[a]

[a]Animal Breeding and Genomics Centre, Animal Sciences Group, Wageningen University and Research Centre, P.O. Box 65,8200 AB Lelystad, the Netherlands

[b] Dipartimento di Agraria – Sezione Scienze Zootecniche, Viale Italia 39, 07100, Sassari. Italy

Corresponding author: Giustino Gaspa, Dipartimento di Agraria – Sezione Scienze Zootecniche, Viale Italia 39, 07100 Sassari, Italy, email: gigaspa@uniss.it , phone +39 079 229309

Abstract

Naturally hornless cattle are called polled. The possibility to introgress the allele responsible (P) for polledness in a cattle population that is more intensively selected for other traits is investigated in this paper. Gene introgression, generally carried out by several steps of backcrossing and selection takes a long time and may lead to unacceptable genetic loss in other traits or inbreeding. The main objective of the current study was to evaluate the use of genomic selection to speed up the introgression of a target allele in a conventional dairy cattle breeding scheme with overlapping generations. A cattle population and a breeding program were simulated and run over a 12 year selection period. Assuming that the polled population was inferior for overall genetic merit , two selection strategies were evaluated to introgress: i) selection on conventional BLUP-EBV (CBLUP); ii) selection on genomic EBV (GEBV) obtained with the genomic relationship matrix used in BLUP (GBLUP). Both selection strategies were applied with (PSEL) and without (NOPSEL) selection for the single polled locus (P). The overall level of genetic merit, the P allele frequency and the inbreeding level (F) in the new born animals were monitored each, as well as the average genetic gain per year of selection (ΔG). The overall genetic level of new born animals was higher for GBLUP compared to CBLUP, with an average ΔG/year of 8.34% (GBLUP) against 7.49% (CBLUP). The PSEL scenario reduced genetic gain, but P allele frequency increased from 0.130 to 0.415 (CBLUP) and from 0.128 to 0.440 (GBLUP) for PSEL, after 12 years of selection. No substantial changes in allele frequency were recorded for NOPSEL scenarios, both for CBLUP and GBLUP breeding schemes. The overall inbreeding rate for GBLUP were 0.28%/y (NOPSEL) and 0.30%/y (PSEL) and for CBLUP 0.52%/y (NOPSEL) and 0.44%/y (PSEL). In conclusion, application of GS to gene introgression helped to speed up the process of introgression of a gene while simultaneously increasing the genetic gain and reducing the inbreeding rate.

Keywords*:* Cattle, marker assisted introgression, genomic selection, polledness, breeding program, dehorning

Introduction

Horned cattle cause management problems and increases veterinary expenses and other economic losses due to injuries during housing or transportation (Prayaga, 2007). Dehorning and disbudding animals are common solutions to overcome these issues, and in Europe 81.5% of dairy cattle and 32.5% of beef cattle are currently dehorned (Cozzi et al. this issue). However, it is suggested that these are painful practices (Stafford and Mellor, 2011), stressful for the animals, and they are perceived by society as not respectful for animal integrity. An alternative might be to breed for naturally hornless cattle called polled cattle. Polledness is based on a single gene, with the polled allele (P) being dominant over the horned allele (p). The gene has been mapped to the centromeric region of bovine chromosome 1 (Georges et al., 1993; Schmutz et al., 1995 Seichter et al. 2012), but, the actual causative mutation of polledness has not yet been identified.

Breeding programs for introgression of the polled allele in beef cattle have been set up in France for Charolais (www.genediffusion.com) and German Fleckvieh (Götze et al, this issue). In Holstein dairy cattle the feasibility to breed polled cows is limited both by lower breeding values for most traits of polled bulls and by the limited number of polled bulls available (Windig and Veerkamp, this issue). A breeding program thus requires introducing the polled allele into the group of cattle while maintaining the high genetic merit of the population.

Gene introgression is generally carried out by several steps of backcrossing and selection. Progeny of cattle from the donor population crossed with cattle from the recipient population carrying the introgressed target allele, are selected for the traits of interest and backcrossed to the recipient population to recover the genome of the original breed (Dekkers and Hospital, 2002; Hospital, 2005). This process is not always straightforward, may take a long time and lead to genetic loss in other

traits. Technological advances in molecular genetics have provided new opportunities for introgressing favourable alleles from inferior populations by using information on genetic markers (Visscher et al., 1996; Wall et al., 2005). Genetic markers can be used to speed up the process of gene introgression directly or to select for (or against) a particular background (Davis and DeNise, 1998). New tools have been developed in animal breeding exploiting the availability of dense marker maps and high-throughput genotyping techniques like genomic selection (GS). GS relies on the segmentation of the genome in thousands of intervals bracketed by contiguous SNPs and SNP effects estimation on the whole genome. Currently up to 800K SNP chip are commercially available for cattle. In the GS framework, the contribution of each SNP to the total genetic variance for a quantitative trait is quantified, and hence the whole genetic variance may be explained by the markers (Goddard and Hayes, 2007).

Genomic selection has been proposed in combination with introgression schemes as a way to incorporate desirable genetic characteristics of a donor line into recipient lines. In aquaculture species, Odegard et al., (2009b) demonstrated through computer simulation that GS is able to produce backcross lines that perform similarly or better than purebred lines within 3 to 5 generations. GS could be used to help to improve the genetic merit of polled cattle by shortening the generation interval and achieving a faster genetic progress compared to selection based on EBV estimation by BLUP.

The main objective of the current study was to evaluate the use of GS to speed up the introgression of the polled allele in a conventional dairy cattle breeding scheme. To study the effect of using GS in a short period of time a population with overlapping generations was simulated under different selection scenarios and for 12 years of selection . Schemes with GS were compared to traditional pedigree based schemes, both for schemes with and without selection for polledness.

Materials And Methods

A cattle population with overlapping generations was simulated and two selection strategies were compared: i) selection on conventionally estimated breeding values (CEBV) with BLUP using the pedigree based numerator relationship matrix (NRM) and ii) genomic EBV (GEBV) obtained with the genomic relationship matrix (GRM). Both selection for polledness (PSEL) and selection without polledness (NOPSEL) were evaluated for CEBV and GEBV. Polledness was simulated following the current genetic model of a single locus with the polled allele (P) being dominant over horned (p*)*. Twenty replicates were performed for each simulated scenario.

A computer simulation program was developed for the evaluation of different selection strategies using FORTRAN 95. The program consisted of five steps (Figure 1) to simulate a cattle breeding program with overlapping generations. First, a base population was set up. The next four steps (breeding value estimation, selection and culling, mating, phenotype generation) were repeated every year for a period of twelve years.

Simulation Program

Population Set Up. First a base population was simulated with an effective population size of 200 animals, with 100 females and 100 males. This structure was kept constant for 1,000 generations in which animals were randomly mated. Two chromosomes of 1 Morgan each and twenty thousand bi-allelic loci were simulated in linkage equilibrium with allele frequencies of 0.5. In the 1000 generations thereafter, each locus had a mutation rate of 2.5×10^{-5} following Habier et al (2007). In total, on average across replicates, ~4,500 loci were segregating in the last generations. The base population for the breeding program itself consisted of genotypes from the last 6 generations, 1,200 individuals in total. From this base population 30 individuals were randomly chosen to be bulls and 570 to be dams to reduce the initial population size to 600 individuals.

Two hundred bi-allelic markers were randomly sampled to be QTLs. The allele substitution effect of each QTL was derived assuming equal contributions of all QTLs to the total additive genetic variance using: $\sigma^2_{g_i} = 2p(1-p)\alpha^2 = \sigma^2_g / n$ where p is the allele frequency at one QTL locus, α is the allele substitution effect, σ^2_g was equal to the simulated variance and n was the number of QTL used (Falconer and Mackay 1996). The simulated additive genetic variance was scaled to obtain a heritability (h^2) of 0.3. The residuals were drawn from a normal distribution with mean 0, and variance $\sigma^2_e = \sigma^2_g \frac{(1-h^2)}{h^2}$. The True Breeding Value (TBV) of individual i was defined as the sum of QTL allelic effects for that individual across the n QTL loci according to

$$TBV_i = \sum_{j=1}^{nQTL} x_{ij} \cdot g_j \qquad [1]$$

where x_{ij} is the number of copies of one of the two alleles that individual i carries at the j-th QTL position and g_j is the effect of the j-th QTL. The phenotypic value was obtained for each animal by adding a random residual $e_i \sim N(0, \sigma^2_e)$ to the TBV.

A locus was chosen among all the neutral markers to mimic the polled locus, taking into account that animals carrying the P allele have lower true breeding values than animals carrying the p allele. The initial P allelic frequency was set to 0.10 (± 0.05) in the base population with the aim to mimic the real frequency in Holstein cattle. All the parameters used in the simulation are summarized in table 1.

Breeding Value Estimation. The selection of individuals to obtain the next generation relied on breeding value estimation, carried out using two models, CBLUP, and GBLUP. In the first model pedigree information only was used to calculate the additive relationship matrix, whereas in the GBLUP approach only marker information was used to compute the genomic relationship matrix. In both cases variance components of random animal effects were re-estimated every year using ASREML software (Gilmour et al. 2009).

CBLUP. The model applied for breeding value estimations, ignoring the marker genotypes, is the classical BLUP animal model as:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zu} + \mathbf{e} \quad [\mathbf{2}]$$

where **y** is a vector of phenotypic values, **1** is a vector of ones, μ is the general mean, **Z** is the incidence matrix that assigns the animals to their phenotypic records, **u** is a vector of random additive genetic effects with distribution ~ N (0, $\mathbf{A}\,\sigma_u^2$) and **e** is a vector of random residuals with distribution ~ N (0, $\mathbf{I}\,\sigma_e^2$), where σ_u^2 and σ_e^2 are the estimates of additive polygenic variance and residual variance, **A** is the additive relationship matrix and **I** is the identity matrix. Individuals in the pedigree and phenotypic records were updated each year as soon as new records were available. The solution of additive animal effects and the variances were estimated together using REML algorithm for every round of selection.

GBLUP. All the genome-wide marker information was used for estimating the genomic breeding values using the genomic relationship matrix approach according to the following model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e} \quad [\mathbf{3}]$$

where **y**, μ , **Z** and **e** are the same as in the CBLUP, whilst **g,** the vector with random animal effect, follow the distribution ~ N (0, $\mathbf{G}\,\sigma_u^2$). The matrix **G** contains the information about the actual fraction of alleles that two individuals share, differently from the **A** matrix which contains the expected fraction of allele shared by two individuals. **G** was calculated as in VanRaden, (2008). $\mathbf{G} = \mathbf{WW}'/2\sum_i p_i(1-p_i)$ where **W** is the centered matrix of SNP genotypes, p_i is the frequency of one of the two allele at i-*th* locus $\sigma_u^2 = 2\sum_i p_i(1-p_i)\,\sigma_g^2$ on n the assumption that σ_g^2 is the genetic variance of marker effects. Variance components (σ_u^2 and σ_e^2) were re-estimated together with animal effects using REML every round of selection. Marker genotypes, individuals in the pedigree and phenotypic records were update as well.

Selection and culling.

The PSEL strategy was carried out grouping the bulls according to their genotype at the polled locus, and ranking them on breeding values. To be selected the bulls needed to carry at least one P allele. When no polled bulls were available, the best horned bulls were used instead. No selection was carried out on the P allele in cows. Cows and bulls were culled at an age greater than 5 and 9 respectively or when their breeding values were lower than available replacements. A further 5% of involuntary culling was also simulated.

Mating and Phenotypes. Selected sires were randomly mated to the selected cows to produce on average the same number of calves each (approximately half male and female). Half of the cows and heifers were mated with young bulls and half with top bulls. According to the marker alleles inherited each calve obtained a TBV. A phenotype was then assigned to the cows born in the previous year (based on their TBV plus a random environmental deviate following equation [1]) adding a new lactation record every year if an animal is within the herd. All calves (both male or female) were kept in the population for one year of breeding values estimation, and the next year the best animals were chosen to be young bulls and heifers based both on their predicted breeding value and polledness for bulls in PSEL scenarios.

Each year, pedigree, phenotypic records and marker genotypes of new born animals were added to the population information to provide the genetic evaluation of all animals.

Breeding Scheme

A dairy breeding scheme and population structure were set up to investigate how genomic selection affects the accuracy of selection (Figure 2). Animals, divided into males and females, were grouped into age class according to the proportion of animal staying alive during the current year t. Females are divided into calves (class of age 0), heifers (class of age 1) and cows (class of age ≥ 2). All female calves are kept in the population (from year t to year t+1) and when they are heifers are selected to either stay in the herd or to be culled. The male line was divided into 4 classes of

individuals: calves, young bulls, waiting bulls and top bulls. Class of age 0 include all the calves born in the current year. Calves to be kept as young bull (class of age 1) are selected according to their breeding values and polledness. Class of age 2-3 included the waiting bulls, which are waiting their genetic evaluation and are not used by the time, whilst the top bulls are grouped in the classes of age >5 (at age 9 are culled).

Figure 3 shows the breeding scheme adopted which contains the number of animals per each category of animal, the number of calves born each year and the modality of mating. At the set up of the population each individual was assigned to an age class, such that each class was equally represented. The selection and culling procedure have been described in the previous sections.

Evaluation of scenarios

The genetic gain, the P allelic frequency and the inbreeding level in the new born animals were monitored over time for all selection criteria. The average inbreeding coefficient (*F*) was calculated for new born animals. Furthermore, the overall inbreeding rate at year *t*=12 (ΔF^i_{1-t}) was calculated for the i-*th* replicate following Falconer and Mackay (1996) as $\Delta F^i_{1-t} = 1 - (1 - F^i_t)^{1/(t-1)}$, where F^i_t was the average inbreeding coefficient in the year *t*=12 for the *i*-th replicate. The maximum among the annual inbreeding rates was also reported (ΔF^i_t), calculated using the formula $\Delta F^i_t = (F^i_t - F^i_{t-1})/(1 - F^i_{t-1})$ where, for the *i*-th replicate, F^i_t and F^i_{t-1} were the average inbreeding coefficients in the years *t* and *t*-1, respectively. The overall ΔG per year was computed regressing the genetic level of the newborn animal, calculated as their average TBV, over time according to $G^i_j = a^i + b^i t + e^i_j$, where, for the *i*-th replicate G^i_j was the (CEBV or GEBV) in the year *j*, b^i the estimated ΔG per year and *t* the number of year of selection. To measure the accuracy of CEBV vs GEBV predictions, for all new born animals was also reported the squared correlation between TBV and CEBV or GEBV respectively.

Results

Simulated Population.

In the base populations on average 4,466 ± 130 markers segregated after 1,000 generations of random mating. This value is the average ± the standard deviation across twenty replicates and all simulated scenarios. The number of segregating markers was similar across different scenarios and replicates (Table 2). The average number of markers per cM was around 22.5. This figure approaches the density for currently available commercial chip sets.

The simulated initial frequency for the P-allele was on average 0.130 and was similar across different scenarios (Table 3). This value was quite close to the target value. Similarly, the values of the simulated heritability of the quantitative trait are in good agreement with the target values (Table 3).

Irrespective of the simulated breeding scheme, the genetic merit constantly increased over the years, although the rate of genetic improvement varied across breeding schemes. As intended, the initial TBV were lower for animals carrying the P allele (-0.071 ± 0.28, for PP animals and -0.056 ± 0.15 for Pp animals respectively) in comparison to pp animals (0.019 ± 0.12).

Effect of different selection strategies on genetic level

Under NOPSEL selection scheme the annual genetic gain was 8.0%/y using CBLUP estimation methods. Under PSEL this figure dropped to 6.9%/y due to the lower breeding values of Polled animals (Table 4). Gains were higher using GBLUP both under NOPSEL (+1.0%) and under PSEL (+0.7%). The cost of introduction of P alleles in the population (PSEL) similar under GBLUP (about -1.3%) and CBLUP (-1.1%) estimation methods (Table 4).

Irrespective to PSEL strategies, the overall genetic gain/year of new born animals was higher under GBLUP compared to CBLUP, with an average value of 8.3%/y (GBLUP) against 7.5%/y (CBLUP).

Genetic gain was relatively low in the first three years. Especially, using CBLUP a sort of lag can be observed. The higher genetic gain under GBLUP only became apparent after 9 years of selection (Figure 4).

Independently from the selection strategy on Polledness adopted, the squared correlation between TBV and GEBV ($r^2_{TBV,GEBV}$) for the new born animals were on average higher (0.705 and 0.731 NOPSEL and PSEL respectively) than for TBV and CEBV ($r^2_{TBV,CEBV}$ 0.209 for both NOPSEL and PSEL strategies, Table 5). These values refer to the approximate reliability of breeding values for the animal at the time of birth for each year of selection. The $r^2_{TBV,GEBV}$ tended to slightly rise along the generations when GBLUP was implemented, due to both the increase of number of genotyped animals (i.e. the reference population) and the frequent re-estimation (each year) of the markers effects (Muir, 2007). Whereas, the $r^2_{TBV,CEBV}$ initially tended to increase, there is no specific trend in later years.

P Allelic frequency

The frequency of P allele tended to be stable in NOPSEL scenarios and increased constantly in the PSEL scenarios. In the case of NOPSEL breeding scheme the P allelic frequency slightly dropped from the initial value of 0.13 to 0.08 (CBLUP) and 0.09 (GBLUP). In some replicates the P allele was lost. (Figure 5). Conversely, the allele frequency raised to 0.41 (CBLUP) and 0.44 (GBLUP) under PSEL. At the phenotypic level around 18% of new born animal are polled in year 12 under NOPSEL compared to about 60% under PSEL. Carriers of the p allele are still common, only 1.5% (NOPSEL) and 12% (PSEL) are homozygous polled after 12 years of selection (Table 4).

Genetic level of Polled vs Horned

All genetic trends were positive especially after the 5th year (Figure 6). The genetic level of PP and Pp genotypes started at negative values because the average TBV was set to 0 at year 1 and polled animals had lower TBVs. Under CBLUP bulls had a lower genetic level after 12 years of selection, than under GBLUP, especially for PP bulls (0.58). Under GBLUP the genetic level of PP animals (0.70) was higher than the genetic merit of horned animals under CBLUP (0.67) after 12 years of selection. Using GBLUP the difference between heterozygous animals and pp animals almost disappeared after twelve years of selection, while the gap between pp and Pp animals remained about the same under CBLUP. The genetic merit of horned (pp) bulls showed the same trend as before switching from conventional to genomic selection breeding scheme.

Inbreeding

The inbreeding levels followed the pattern of genetic gain in all four simulated scenarios. Under NOPSEL and CBLUP $\Delta F_{1\text{-}12}$ was 0.52%/y whereas this figure dropped to 0.44%/y in the case of PSEL. Under GBLUP inbreeding rates were lower both for NOPSEL (0.28%/y) and PSEL (0.30%/y) Inbreeding rates under CBLUP varied more compared to GBLUP (Table 4). Consequently, the maximum among annual inbreeding rate (Max ΔF) was higher for CBLUP (0.93% (NOPSEL) and 0.94% (PSEL)) compared to GBLUP (0.61%/y and 0.59%/y for NOPSEL and PSEL respectively). The difference in inbreeding level between CBLUP and GBLUP only becomes apparent after 5 years of selection (Figure 7).

Discussion

This investigation is based on the idea that introgression of the P allele is a valuable option for breeding polled sires of high genetic merit. Thanks to the advance in molecular techniques and

288 the drop of genotyping costs, the use of genomic selection will allow to increase the accuracy of
289 selection, shorten the generation interval and increase the genetic level of polled animals.
290
291 **Effect of different selection strategies on genetic level**
292 The differences observed in genetic level between GBLUP and CBLUP show the potential advantage
293 of using GS when introgressing a target allele in a population, as also shown by Odegard et al.,
294 (2009a) who used GS to speed up the introgression of a major QTL from a donor line of lower genetic
295 level to a recipient line of higher genetic merit, and where conventional EBV selection was inefficient
296 to preserve the target QTL. Because of a longer generation interval and generally complex population
297 structures of livestock populations only a few examples of successful marker assisted introgression
298 (MAI) programs have been reported in literature (Hanset et al., 1995; Yancovich et al., 1996). With
299 MAI knowledge in advance of the location and the effect of QTL is needed, whilst with genomic
300 selection the introgression and the genetic improvement of selected candidates is done simultaneously
301 and any favorable allele from the "donor" population may be introgressed (Odegard et al., 2009a). In
302 case of absence of selection the recovery of recipient genome surrounding the introgressed gene
303 occurs at lower rate than the case of continue selection and it is function of intensity of selection
304 adopted. Results from Hospital et al. (1992) on MAI showed, even with very few marker per
305 chromosome, that the response of selection is different on carrier and not carrier chromosome, being
306 lower for the former. In the current study only two chromosome were simulate (one carrier and one
307 not) for computational time constrains. Moreover thousand of marker were simulated here.
308 Simulation of the whole cattle genome of 30 Morgan should give a precise estimation of that effect.
309 However, since genomic selection potentially allows to capture all the genetic variance of the trait
310 due to QTL in LD with markers, a better estimation of effect of this segment of DNA might result in
311 higher response of selection also for non carrier chromosome.
312

313 Results from this simulation show how GS increases the accuracy of breeding values of new
314 born animals irrespective whether the breeding scheme for polledness is adopted or not. Previous
315 simulation studies showed the potential of GS to increase the accuracy of selection and the genetic
316 gain in dairy cattle (Schaeffer, 2006; de Roos et al., 2011; Lillehammer et al., 2011; Buch et al.,
317 2012). The increase in accuracy due to GS compared to conventional selection is associated with the
318 selective pressure on mendelian sampling terms (Daetwyler et al., 2007). The level of GEBV accuracy
319 in the present study increased in the first 5 year and did not vary further. Several authors reported the
320 decrease of accuracy over time due to recombination (Habier et al., 2007; Muir, 2007). In the present
321 study the continued increase of the accuracy along the generations is due both to the yearly re-
322 estimation of marker effects and the increasing number of generations with phenotypic and genomic
323 information that is included in the reference population. These factors as cause of variation in GEBV
324 accuracy have been reported before both in simulated and field data (Habier et al., 2007; Muir, 2007
325 ; Liu et al., 2011).
326
327 **P allelic frequency**
328 The frequency of the P allele is about constant when no selection on P is performed but
329 increased under selection. There is no clear difference between GBLUP and CBLUP scenario apart
330 from sampling error which was higher for CBLUP. The increase of P allelic frequency is moderate
331 because we choose to select animals with at least one P allele as parents, rather than PP animals if the
332 latter were of lower genetic merit. The allelic frequency reached in the present study after twelve
333 years of relatively mild selection is in agreement to a program of introgression of favorable alleles
334 for sensitivity at scrapie in sheep through computer simulation (Windig et al., 2004). Such mild
335 selection may represent a good compromise between maximization of genetic gain per year, and
336 maintaining an acceptable inbreeding level. A steep increase of the P allelic frequency can be reached
337 using exclusively homozygous animal at P locus and should be joined to a genomic selection breeding
338 scheme to control the inbreeding level in a more effective way. On the other hand, focusing solely on

polled status, the performance in economically important traits may be compromised in a conventional EBV breeding program. Depending on the breeding strategy the transition to a polled herd may take a long period of time. Shortening the generation interval due to the application of genomic selection scheme favors the introgression. Furthermore, even if not considered in the present study, putting priority on culling phenotypically horned cows will reduce the p allele frequency even faster.

Genetic level of Polled *vs* Horned.

The genetic level of polled (both P/- or P/P) animals increased constantly during the selection program for polledness. This increase of the breeding values of Polled animals was especially clear in the later year of selection when the genomic selection became effective. The use of genomic selection in GBLUP scenarios allows to have an extra gain both for P/- and PP new born animals in comparison to CBLUP scenarios, due both to the higher accuracy of breeding values and shortened generation interval. The usefulness of application of genomic selection breeding scheme was also shown (extra ΔG from 28% up to 108% depending on the % of young bulls used) in breeding scheme simulation studies which used both deterministic and stochastic approaches (König and Swalve 2009; Pryce et al., 2010; de Roos et al., 2011).

Inbreeding rate

In the present study the annual inbreeding rate for GBLUP breeding scheme were lower than conventional CBLUP breeding program. A similar example of the reduction of inbreeding level by application of marker assisted selection was reported in a simulation study by Pedersen et al., (2009). The lower inbreeding rates of GBLUP could be due to a better estimation of Mendelian sampling terms and minimization of co-selection of sibs as shown by Daetwyler et al. (2007), and selection of younger animals as also shown by Lillehammer et al., (2011) but opposite to the finding of de Roos et al., (2011) who in their simulation study found that the shortening of the generation interval doubled

365 the genetic gain/year at the same inbreeding rate per generation, at cost of a higher inbreeding rate
366 per year. In our study the selection strategy affected the overall inbreeding rates. Although, we
367 computed the inbreeding at pedigree level, several authors stated that in GS program it is better to
368 control the rate of inbreeding at genomic level (Sonesson et al., 2012; Pryce et al. 2013). Rates were
369 slightly higher in case of PSEL scenario in comparison to NOPSEL for GBLUP. Conversely, in the
370 case of CBLUP the higher figure is recorded for NOPSEL.
371

372 ## Conclusion

373

374 Currently in Holstein cattle few bulls are available with the polled gene and they have on
375 average low breeding values for the other breeding goal traits. Selection for polledness will lead to
376 an (initial) decrease in the other breeding goal traits that could be contained using GS. The achieved
377 accuracy of genomic selection allows to increase the breeding values of polled sires faster than in a
378 conventional breeding scheme. The simultaneous increase of the frequency of the P allele, and
379 breeding values of Polled animals via genomic selection resulted in lower inbreeding rates compared
380 to conventional EBV breeding scheme.
381

382 ## Acknowledgments

383 The authors are grateful for discussions with Han Mulder (Animal Breeding and Genomic
384 Centre, Wageningen University). The study was also performed as part of the RobustMilk and
385 SABRETRAIN projects. The RobustMilk project is financially supported by the European
386 Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708.
387 The first author was funded by the SABRETRAIN Marie Curie Host Fellowships for Early Stage
388 Research Training, as part of the 6th Framework Programme of the European Commission. This

publication represents the views of the authors, not the European Commission, and the Commission is not liable for any use that may be made of the information

Conflict of interest statement.

I have no conflicts of interest to declare.

References

Buch, L.H., Sorensen, M.K., Berg, P., Pedersen, L.D., Sorensen, A.C., 2012. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. Journal of Animal Breeding and Genetics 129, 138-151.

Daetwyler, H.D., Villanueva, B., Bijma, P., Woolliams, J.A., 2007. Inbreeding in genome-wide selection. Journal of Animal Breeding and Genetics 124, 369-376.

Davis, G.P., DeNise, S.K., 1998. The impact of genetic markers on selection. J. Anim Sci. 76, 2331-2339.

de Roos, A.P.W., Schrooten, C., Veerkamp, R.F., van Arendonk, J.A.M., 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. Journal of Dairy Science 94, 1559-1567.

Dekkers, J.C.M., Hospital, F., 2002. The use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genetics 3, 22-32.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Georges, M., Dietz, A.B., Mishra, A., Nielsen, D., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X.Y., Leipold, H., Womack, J.E., Lathrop, M., 1993. Microsatellite Mapping of the Gene Causing Weaver Disease in Cattle Will Allow the Study of an Associated Quantitative Trait Locus. P Natl Acad Sci USA 90, 1058-1062.

Goddard, M.E., Hayes, B.J., 2007. Genomic selection. J Anim Breed Genet 124, 323-330.

Habier, D., Fernando, R.L., Dekkers, J.C.M., 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177, 2389-2397.

Hanset, R., Dasnois, C., Scalais, S., Michaux, C., Grobet, L., 1995. Introgression into the Pietrain Genome of the Normal Allele at the Locus for Halothane Sensitivity. Genet. Sel. Evol. 27, 77-88.

Hospital, F., Chevalet, C., Mulsant, P., 1992. Using Markers in Gene Introgression Breeding Programs. Genetics 132, 1199-1210.

Hospital, F., 2005. Selection in backcross programmes. Philos T R Soc B 360, 1503-1511.

Konig, S., Simianer, H., Willam, A., 2009. Economic evaluation of genomic breeding programs. Journal of Dairy Science 92, 382-391.

Lillehammer, M., Meuwissen, T.H., Sonesson, A.K., 2011. A comparison of dairy cattle breeding designs that use genomic selection. J Dairy Sci 94, 493-500.

Liu, Z.T., Seefried, F.R., Reinhardt, F., Rensing, S., Thaller, G., Reents, R., 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet. Sel. Evol. 43, -.

Muir, W.M., 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. Journal of Animal Breeding and Genetics 124, 342-355.

Odegard, J., Sonesson, A.K., Yazdi, M.H., Meuwissen, T.H.E., 2009a. Introgression of a major QTL from an inferior into a superior population using genomic selection. Genet. Sel. Evol. 41, -.

Odegard, J., Yazdi, M.H., Sonesson, A.K., Meutwissen, T.H.E., 2009b. Incorporating Desirable Genetic Characteristics From an Inferior Into a Superior Population Using Genomic Selection. Genetics 181, 737-745.

Pedersen, L.D., Sorensen, A.C., Berg, P., 2009. Marker-assisted selection can reduce true as well as pedigree-estimated inbreeding. Journal of Dairy Science 92, 2214-2223.

Prayaga, K.C., 2007. Genetic options to replace dehorning in beef cattle - a review. Aust J Agr Res 58, 1-8.

444 Pryce, J.E., Goddard, M.E., Raadsma, H.W., Hayes, B.J., 2010. Deterministic models of breeding
445 scheme designs that incorporate genomic selection. Journal of Dairy Science 93, 5455-5466.
446 Pryce, J., Hayes, B., Goddard, M., 2012. Novel strategies to minimize progeny inbreeding while
447 maximizing genetic gain using genomic information. J Dairy Sci 95, 377 - 388.
448 Seichter, D., Russ, I., Rothammer, S., Eder, J., Forster, M., Medugorac, I., 2012. SNP-based
449 association mapping of the polled gene in divergent cattle breeds. Anim Genet 43, 595-598.
450 Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal
451 Breeding and Genetics 123, 218-223.
452 Schmutz, S.M., Marquess, F.L.S., Berryere, T.G., Moker, J.S., 1995. DNA Marker-Assisted Selection
453 of the Polled Condition in Charolais Cattle. Mamm Genome 6, 710-713.
454 Sonesson, A., Woolliams, J., Meuwissen, T., 2012. Genomic selection requires genomic control of
455 inbreeding. Genet Sel Evol 44, 27.
456 Stafford, K.J., Mellor, D.J., 2011. Addressing the pain associated with disbudding and dehorning in
457 cattle. Appl Anim Behav Sci 135, 226-231.
458 VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. Journal of Dairy
459 Science 91, 4414-4423.
460 Visscher, P.M., Haley, C.S., Thompson, R., 1996. Marker-assisted introgression in backcross
461 breeding programs. Genetics 144, 1923-1932.
462 Wall, E., Visscher, P.M., Hospital, F., Woolliams, J.A., 2005. Genomic contributions in livestock
463 gene introgression programmes. Genet. Sel. Evol. 37, 291-313.
464 Windig, J.J., Eding, H., Moll, L., Kaal, L., 2004. Effects on inbreeding of different strategies aimed
465 at eliminating scrapie sensitivity alleles in rare sheep breeds in The Netherlands. Anim Sci 79,
466 11-20.
467 Yancovich, A., Levin, I., Cahaner, A., Hillel, J., 1996. Introgression of the avian naked neck gene
468 assisted by DNA fingerprints. Anim Genet 27, 149-155.
469

Table 1. Parameters used to define the simulated population and the breeding scheme.

Items	Values
Genome length	200 cM
Number of chromosome	2
Average marker per cM	22.5
Number of QTL simulated	200
Distribution of QTL effect	$N \sim (0, \sigma^2_{gi})$
Number of generation of RM	1,000
Mutation rate of QTL or marker locus	2.5×10^{-5}
Recombination Rate (θ)	$\theta = f(d)$ Haldane
Initial frequency of P allele	0.10 ± 0.05
Initial Population size	100 males/100 females
Number of animal at gen 1000	600 males/600 females
Number of animal at gen 1001	30 males/570 females
Fertility	1
Involuntary culling	0.05
Number of calves born each year	475
Number of progeny/cows	1
Progeny per sire	~ 30
Year of selection	12
Heritability	0.3
Number of Replicates performed	20

473 **Table 2**. Number of segregating loci after 1000 generation of random mating in the base population

474 (1 Morgan) and average number of marker simulated ± standard deviation across 20 replicates.

	Estimation Method and scenarios			
	CBLUP		GBLUP	
Chromosome	NOPSEL	PSEL	NOPSEL	PSEL
1	2,268 ± 141	2,246 ± 146	2,238 ± 139	2,236 ± 112
2	2,181 ± 103	2,243 ± 150	2,232 ± 136	2,222 ± 142
total	4,449 ± 122	4,489 ± 148	4,470 ± 137	4,459 ± 127

475

476

477

Table 3. Expected P allelic frequency, E(freq P), and simulated P allelic frequency, Sim(Freq P), in the base population. Expected heritability, E(h^2), actual simulated h^2 in the base population and estimated h^2 using CBLUP or GBLUP in both scenarios of polled selection (PSEL or NOPSEL) across 20 replicates.

Scenario	E(freq P) ± sd	Sim(Freq P) ± sd	E(h^2)	Simulated h^2 ± sd	Estimated h^2 ± se
CBLUP-NOPSEL	0.100 ± 0.05	0.130 ± 0.030	0.300	0.295 ± 0.047	0.290 ± 0.024
CBLUP-PSEL	0.100 ± 0.05	0.131 ± 0.030	0.300	0.307 ± 0.048	0.300 ± 0.024
GBLUP-NOPSEL	0.100 ± 0.05	0.128 ± 0.030	0.300	0.280 ± 0.039	0.307 ± 0.030
GBLUP-PSEL	0.100 ± 0.05	0.129 ± 0.029	0.300	0.302 ± 0.056	0.293 ± 0.031

Table 4. Genetic gain/year ± sd ($\Delta G_{1\text{-}12}$) for the simulated trait, P allelic frequency ± sd in new born animal after 12 years of selection, (P_{12}), Genotypic frequency (± sd) of homozygous Polled and heterozygous Polled in the base population (P/P_0 and $P/-_0$) and after 12 years of selection (P/P_{12} and $P/-_{12}$). All values are averaged across 20 replicates..

	Scenario			
	CBLUP		GBLUP	
Items	NOPSEL	PSEL	NOPSEL	PSEL
$\Delta G_{1\text{-}12}$ /year (%)	8.00 ± 1.22	6.94 ± 1.09	9.00 ± 1.41	7.68 ± 1.10
Freq P_{12} newborn	0.085 ± 0.060	0.415 ± 0.095	0.092 ± 0.065	0.440 ± 0.088
Freq P/P_0	0.025 ± 0.015	0.026 ± 0.012	0.020 ± 0.013	0.030 ± 0.015
Freq P/P_{12}	0.015 ± 0.032	0.117 ± 0.056	0.015 ± 0.019	0.110 ± 0.058
Freq $P/-_0$	0.274 ± 0.113	0.289 ± 0.064	0.270 ± 0.087	0.292 ± 0.082
Freq $P/-_{12}$	0.173 ± 0.145	0.618 ± 0.089	0.196 ± 0.122	0.602 ± 0.129

Table 5. Squared correlation (± sd) of CEBV (or GEBV) on True Breeding Vaules ($r^2_{EBV\text{-}TBV}$) in 4 different scenarios calculated for new born animals in the reference year and averaged across 20 replicates.

	$r^2_{EBV\text{-}TBV}$			
	CBLUP		GBLUP	
Year	NOPSEL	PSEL	NOPSEL	NOPSEL
1	0.077 ± 0.126	0.086 ± 0.113	0.641 ± 0.040	0.594 ± 0.060
2	0.240 ± 0.086	0.230 ± 0.068	0.668 ± 0.046	0.662 ± 0.046
3	0.236 ± 0.079	0.249 ± 0.081	0.715 ± 0.033	0.690 ± 0.053
4	0.247 ± 0.085	0.255 ± 0.070	0.714 ± 0.035	0.719 ± 0.043
5	0.250 ± 0.048	0.242 ± 0.054	0.744 ± 0.026	0.740 ± 0.040
6	0.255 ± 0.079	0.261 ± 0.066	0.756 ± 0.035	0.758 ± 0.032
7	0.258 ± 0.042	0.239 ± 0.066	0.765 ± 0.031	0.738 ± 0.039
8	0.201 ± 0.043	0.185 ± 0.047	0.747 ± 0.035	0.697 ± 0.060
9	0.189 ± 0.080	0.191 ± 0.071	0.741 ± 0.043	0.706 ± 0.065
10	0.183 ± 0.078	0.200 ± 0.088	0.767 ± 0.032	0.711 ± 0.069
11	0.191 ± 0.064	0.173 ± 0.067	0.767 ± 0.043	0.729 ± 0.093
12	0.185 ± 0.089	0.197 ± 0.056	0.754 ± 0.041	0.720 ± 0.080
Average ± sd	0.209 ± 0.075	0.209 ± 0.071	0.731 ± 0.037	0.705 ± 0.057

Figures

Figure 1. Five step outline of simulated breeding program. Each step represent a module of the simulated breeding program.

Figure 2. Flow of animals from year *t* to year *t+1.* The population is split in male and female and all selected animals in the current year are signed by (s). Different flows are indicated with 3 type of arrow: flow of animal born in the previous year or animals already in the population (solid arrow) flow of animal born during the year (dotted arrow), flow of animal culled during the year (dashed row) .

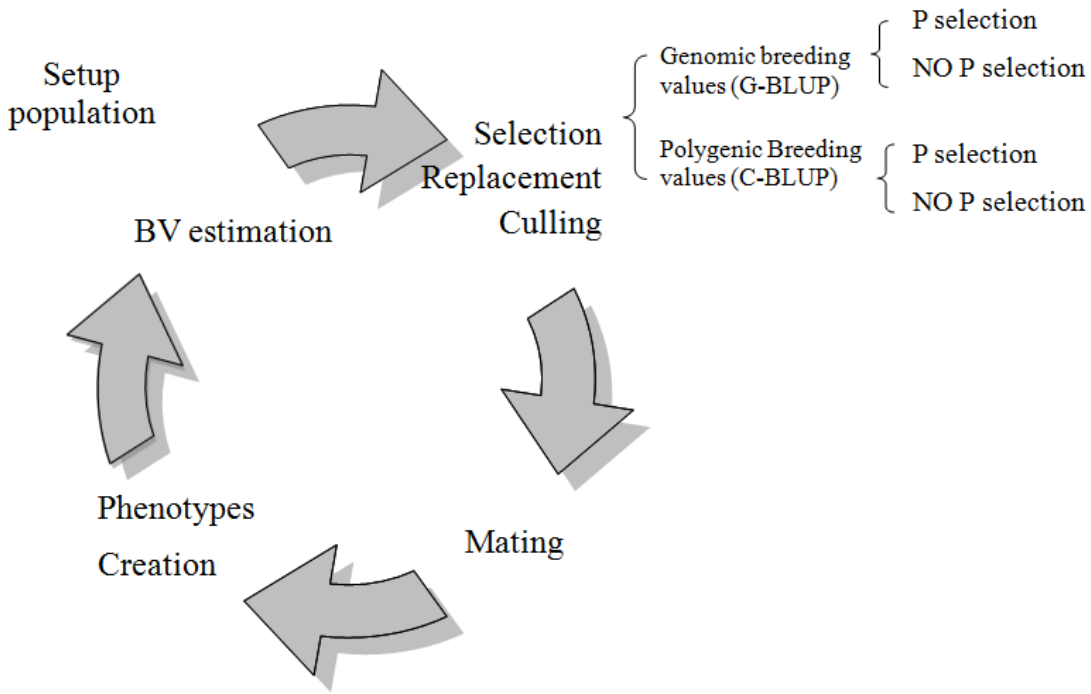
Figure 3. Structure of the breeding scheme

Figure 4. Average breeding values (GEBV or CEBV) of new born animals in each year averaged across 20 replicates in four analyzed scenarios (bar represent standard deviation): polled selection using either CBLUP (CBLUP-PSEL) or GBLUP (GBLUP-PSEL); No selection on polledness using both CBLUP (CBLUP-NOPSEL) and GBLUP (GBLUP-NOPSEL).

Figure 5. Pattern of P allelic frequency (each gray line is a replicate, whereas the black line is the average) in new born animals along 12 years of selection across 20 replicates in: NOPSEL (a, b) and PSEL scenarios (c, d) respectively, in case of either CBLUP (a, c) or GBLUP (b, d) breeding value estimation method used.

Figure 6. Genetic level of new born animals according to the breeding values estimation method (GBLUP *vs* CBLUP) and genotypes at polled locus (polled-PP and Pp; horned-pp) in PSEL scenarios. Values are the average of 20 replicates.

Figure 7. Average inbreeding level (%) measured by mean of F coefficient across 20 replicates (bar indicate standard deviation) either in case of no selection on polled locus (a) or in case of selection on polled locus (b) using both CBLUP or GBLUP breeding value estimation methods.

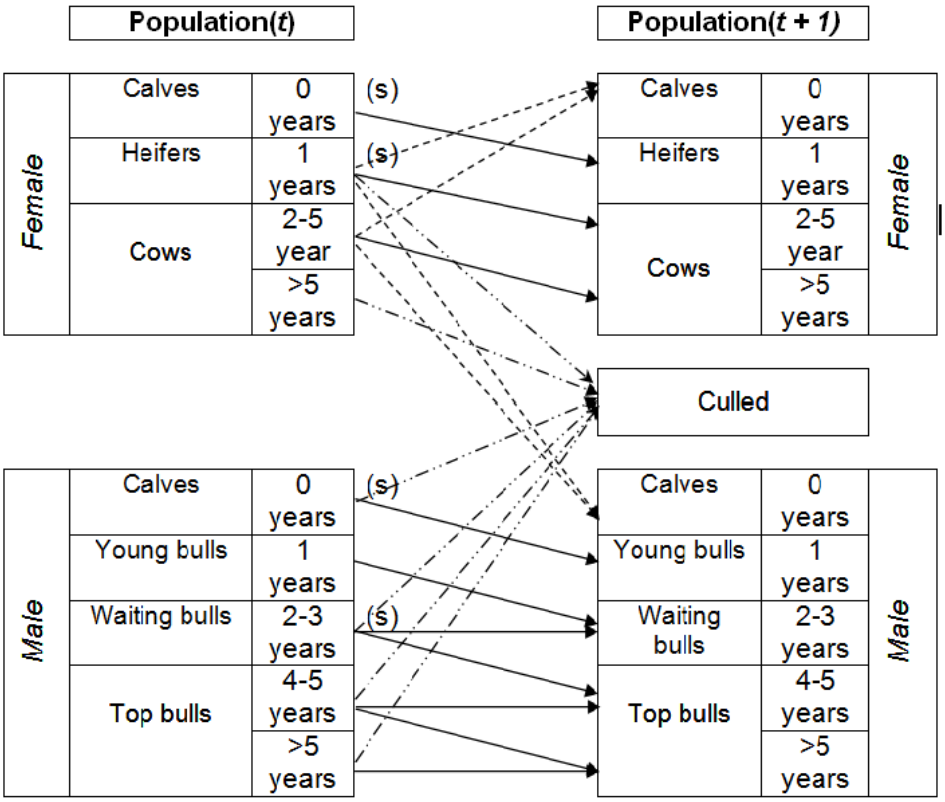


520

521

522 **Figure 1.**

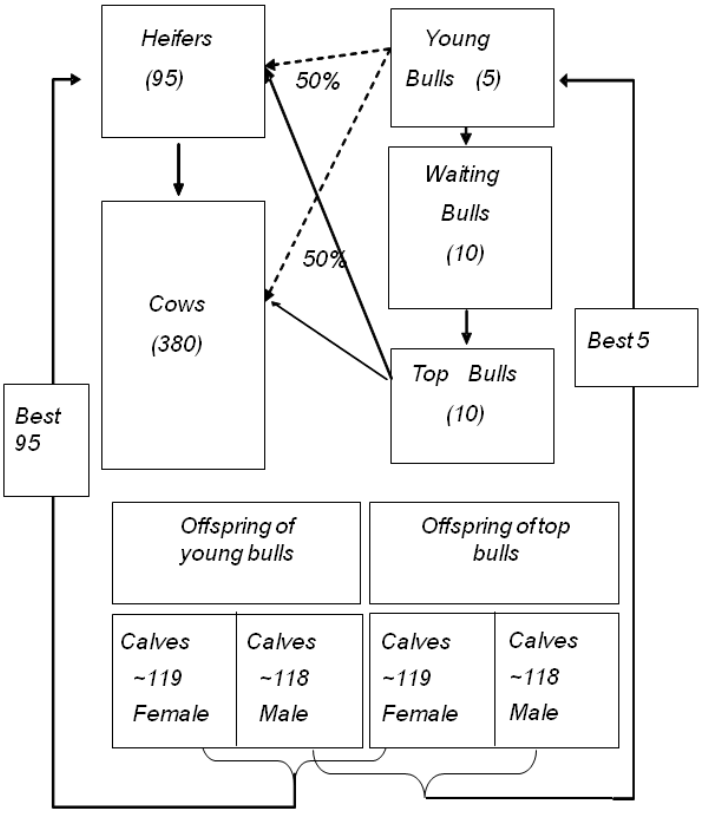
523



524

525 **Figure 2.**

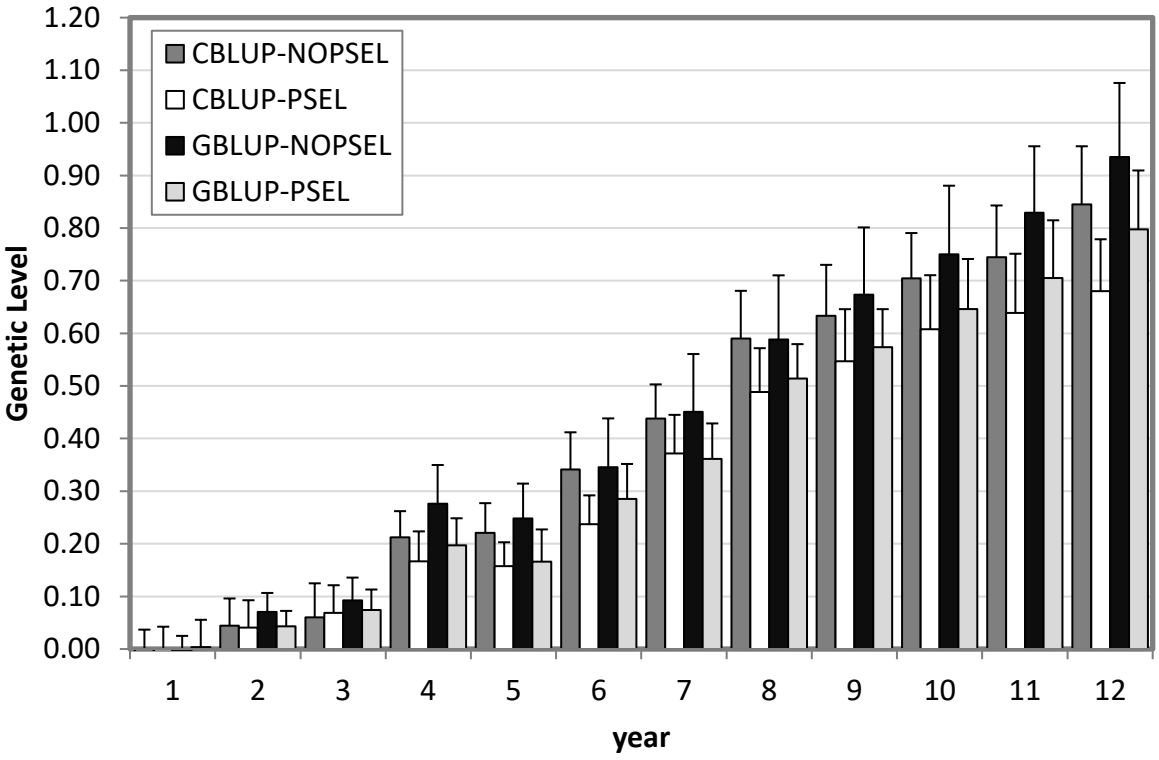
526



527

528 **Figure 3**

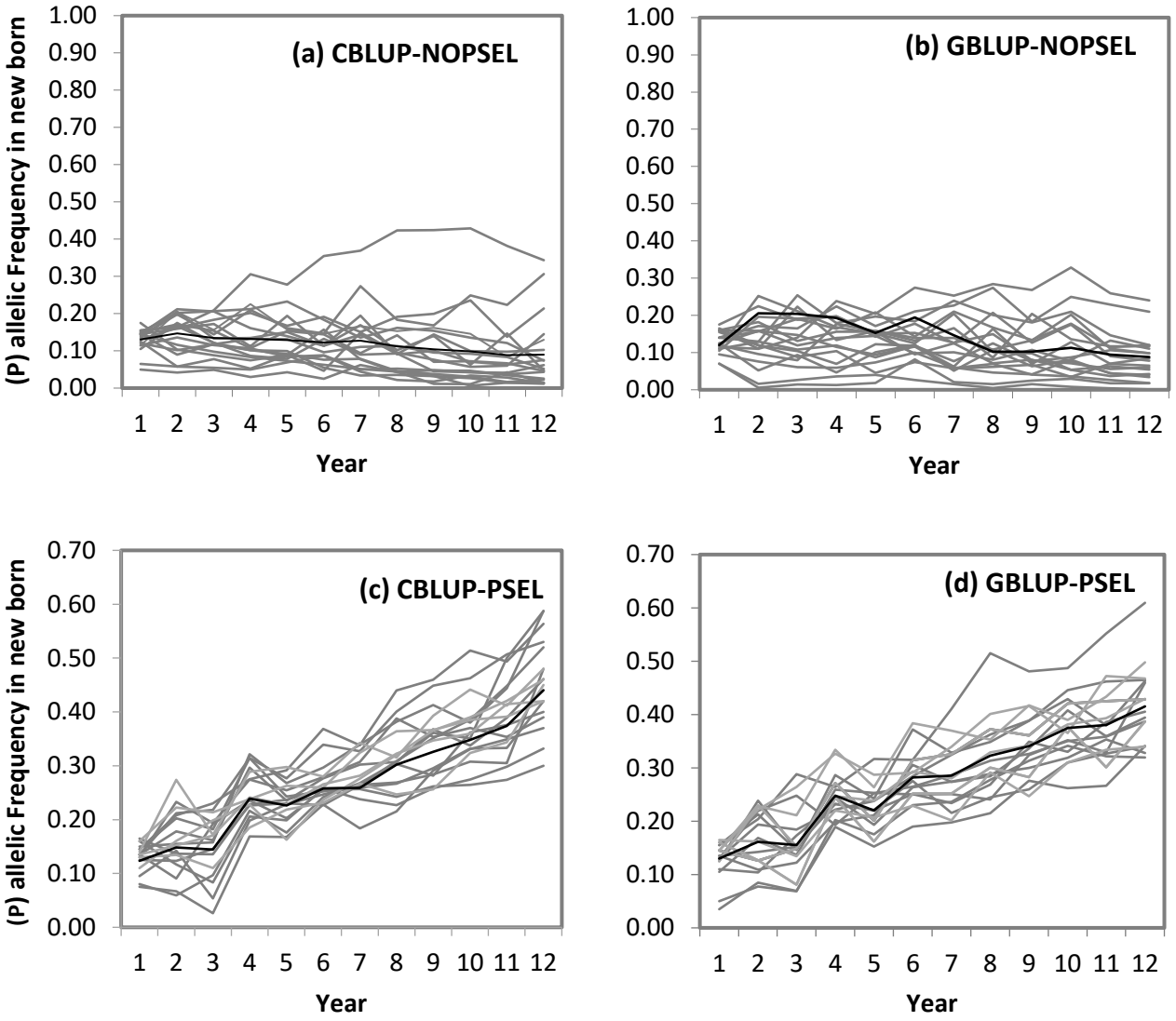
529



530

531 **Figure 4.**

532

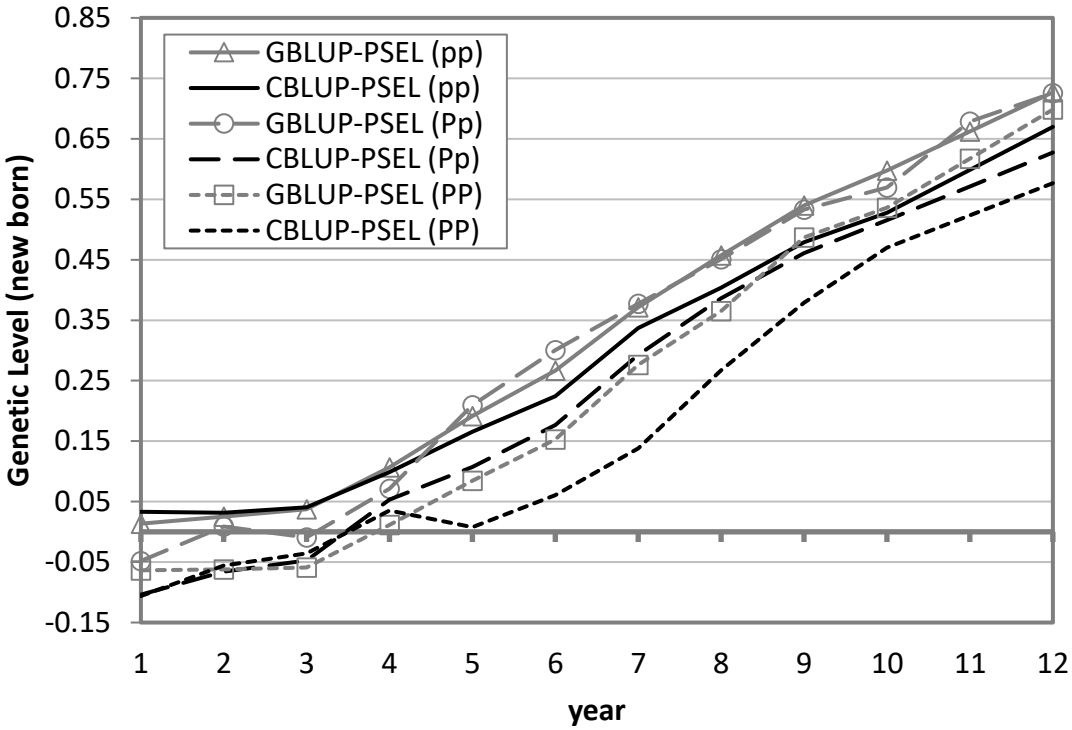


533

534
535

536 **Figure 5.**

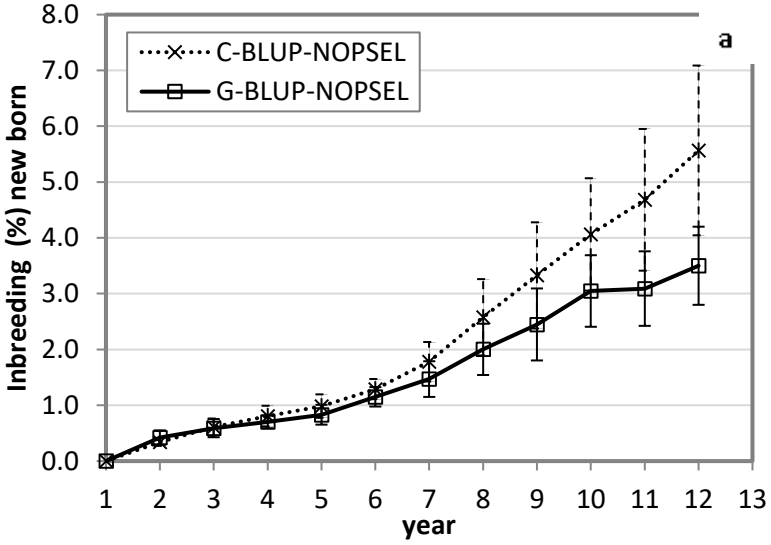
537



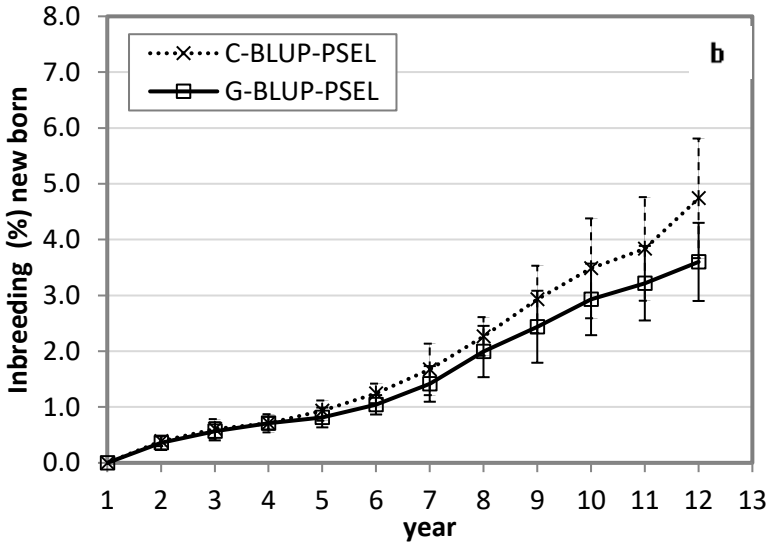
538

539 **Figure 6.**

540



541



542

543

544 **Figure 7.**

545